Deliverable D 3.2

WP3 Report on AI approaches and models

Project acronym:	RAILS
Starting date:	01/12/2019
Duration (in months):	36
Call (part) identifier:	H2020-S2R-OC-IPX-01-2019
Grant agreement no:	881782
Due date of deliverable:	Month 31
Actual submission date:	July 31^{st} 2022
Responsible/Author:	Francesco Flammini (LNU)
Dissemination level:	Public
Status:	Issued

Reviewed: yes

Document history		
Revision	*Date*	*Description*
1	June 28th 2022	First issue for internal review
2	July 31st 2022	Second issue after internal review
3	March 1st 2023	Third issue after JU review

Report contributors		
Name	Beneficiary Short Name	Details of contribution
Francesco Flammini	LNU	Coordinator
Lorenzo De Donato	CINI	Contributor
Zhiyuan Lin	LEEDS	Contributor
Ruifan Tang	LEEDS	Contributor

Advisory Board Reviewers	
Name	*Company or Institution*
Alice Consilvio	University of Genoa, Italy
Nunzia Visciano	Syenmaint, Italy

Funding

This project has received funding from the Shift2Rail Joint Undertaking under the European Union's Horizon 2020 research and innovation programme under grant agreement n. 881782 Rails. The JU receives support from the European Union's Horizon 2020 research and innovation program and the Shift2Rail JU members other than the Union.

Contents

Executive Summary

This Deliverable leverages the results achieved in Deliverable D3.1 to address AI approaches customised to the rail sector. On the basis of the pilot cases studies selected in Deliverable D3.1, i.e., smart maintenance at level crossings, and predictive maintenance of railway assets using digital twin models, specific proofs-of-concept are provided to investigate the adoption of AI methods for smarter railway maintenance and defect detection.

Specifically, the report addresses the following themes for each case study: i) a description of the background to identify the main objectives and the open issues; ii) the identification of proper research questions; iii) the selection of safety and performance indicators to be used as evaluation criteria; iv) the analysis of technology and methodological alternatives; v) a high-level description of the selected AI approach and its possible adaptation to the railways; vi) the identification of the main technological and architectural issues to be considered for future implementations.

The proposed proofs-of-concept are intended to give a first insight towards the definition of qualitative and technology roadmaps which could lead to the deployment of AI applications aiming at enhancing rail maintenance and defect detection.

Abbreviations and acronyms

Abbreviations / Acronyms	Description
AI	Artificial Intelligence
BA	Barrier Analysis
CBM	Condition-based Maintenance
CNN	Convolutional Neural Network
DL	Deep Learning
GE	Generic Alarm
GPS	Global Positioning System
GTA V	Grand Theft Auto V
KPIs	Key Performance Indicators
IoT	Internet of Things
LC	Level Crossing
ML	Machine Learning
NA	No Alarm
PoC	Proof of Concept
R-CNN	Region-based CNN
RQ	Research Question
RUL	Remaining Useful Life
WB	Warning Bell
WBD	Warning Bell Detection
WL	Warning Light
WLD	Warning Light Detection

1. Background

The present document constitutes the Deliverable D3.2 “WP3 Report on AI approaches and models” of the Shift2Rail JU project “Roadmaps for AI integration in the Rail Sector” (RAILS). The project is in the framework of Shift2Rail’s Innovation Programme IPX. As such, RAILS does not focus on a specific domain, nor does it directly contribute to specific Technical Demonstrators but contributes to Disruptive Innovation and Exploratory Research in the field of Artificial Intelligence within the Shift2Rail Innovation Programme. The successor of the Shift2Rail Joint Undertaking is currently the “Europe’s Rail Joint Undertaking” (EU-Rail) established by Council Regulation (EU) 2021/2085 of 19 November 2021. The RAILS workpackage WP3 investigates the adoption of learning techniques and other AI methods for enhanced rail safety and automation. The present Deliverable describes the work carried out in task 3.2 (Development of AI approaches and models) whose aim is to develop AI approaches according to the results and choices reported in D3.1:

- Directions for transferability, including among identified research lines a) Remaining Useful Life (RUL) estimation through non-intrusive sensors; b) AI-aided fault diagnosis and prognosis through Digital Twins and IoT;
- Pilot case studies, specifically addressing anomaly detection at Level Crossings (LC) and intelligent predictive maintenance for rolling stock at railway station depots and workshops by using Digital Twins and machine learning.

The two pilot case studies mentioned above aim at providing the context for proofs-of-concept. In the development of the proofs-of-concept, the transferability of concepts and techniques from other sectors are considered, since smart maintenance and defect detection are common in several other domains. The scope of this deliverable to prepare the ground for the actual experimentation, which will be performed in the continuation of the project in order to highlight the main challenges, obstacles, opportunities, and potential of artificial intelligence and machine learning when applied to railway maintenance applications. Those results will be used to define the future roadmaps for effective AI integration in the railway maintenance sector.

2. Objective

The objective of the RAILS workpackages WP2, WP3, and WP4 is to define pilot case studies and develop proofs-of-concept leading to a technology roadmapping for an effective pick-up of AI in the rail sector. Specifically, the project activities reported in this document and in the next Deliverable D3.3 address Objectives 4 and 5 of the RAILS project:

- *Development of methodological and experimental proofs-of-concept*;
- *Development of Benchmarks, Models and Simulations*.

This deliverable takes its input from Deliverable D3.1 ("Report on case studies and analysis of transferability from other sectors") and formalises the pilot case studies identified in D3.1, so providing feasibility studies for the adoption of AI and related techniques in the area of rail maintenance and defect detection.
More specifically, we selected two relevant case studies in order to investigate real-world requirements, challenges and perspectives when applying artificial intelligence and machine learning to smart maintenance in the railway domain.

The first case study addresses railway level crossings, which are among the most critical railway assets in terms of accidents and maintenance, due to the difficult accessibility and high costs in regular human monitoring and supervision. In such a scenario, it would be very important to leverage artificial vision for detecting anomalies in warning lights and barrier movement, as well as intelligent audio recognition to detect malfunctions in warning bells.

The second case study relates to predictive maintenance for rolling stock in railway station depots and workshops by using artificial intelligence and digital twins. Predictive maintenance is one of the most successful applications of machine learning in many domains, including railways [1]. It is used for detecting defects and anomalies by inspection mainly based on artificial vision. Detailed models based on digital twins and machine learning allow accurate defect detection and fault prediction.

Both case studies share several challenges such as the availability of appropriate datasets and achievable accuracy in order for those AI systems to be feasible and trusted in real-world applications.

3. Introduction

This document provides a detailed description of the AI solutions and approaches for addressing the problems and the challenges posed by the reference applications, the related models and metrics, and the technological and operational issues. Hence, this deliverable reports on the possible technological and methodological solutions, alternatives and criticalities to be addressed by subsequent implementations.

The heart of this document consists of two main chapters, addressing the aforementioned issues for the two pilot case studies: Chapter 4 is devoted to Smart Maintenance at Level Crossings, and Chapter 5 deals with AI-based Predictive Maintenance for Rolling Stock.

The two chapters share the same structure: a brief presentation of the scope and challenges of the selected case study (Sections 4.1 and 5.1), the specific objectives of related proofs-of-concept (Sections 4.3 and 5.3), a description of the proposed approach and exploitable tools (Sections 4.4 and 5.4), datasets (Sections 4.5 and 5.6), a presentation of the AI techniques being exploited in the development of the proof-of-concept (Sections 4.6 and 5.5), and a discussion about expected results and possible criticalities (Sections 4.7 and 5.7).

The implementation and experimentation activities are ongoing according to the objectives, approaches and techniques reported in this document. Details and results of these activities are the objects of the next Deliverable D3.3 ("Report on experimentation, analysis and discussion of results").

4. Smart Maintenance at Level Crossings

4.1. Introduction

As introduced in our previous deliverable D3.1 [2], Level Crossings (LCs) represent a class of railway assets that raise several concerns in terms of both safety and maintainability. Therefore, as also suggested by the Advisory Board, they are a relevant case study to propose benchmarks, proofs-of-concept, and roadmaps.

One of the latest safety reports from the European Railway Agency for Railways (ERA) [3] states that, at the European Union level, fatalities at Level Crossings account for about 30% of all the fatalities registered in railways scenarios in the reference period 2015-2019. This statistic has also been confirmed in 2020 ($\sim$ 31%)[1]. Hence, these railway assets are so sensitive from safety perspective that actions are required to make them safer. Attempts have been made to replace level crossings with subways or bridges [4], nevertheless, this replacement process is too expensive given the huge number of LCs (about 105 thousand within EU countries). Therefore, other short-term interventions should be taken into account such as those proposed within the SAFER-LC project [5] (concerning new warning signalling systems, speed bumps, but also detection and risk evaluation systems based on machine learning) or those reviewed in [6] (mostly based on IoT and train-to-wayside communication). Also, reference [7] gave a complete review of suitable obstacle detection technologies and their associated algorithms that can be exploited to support risk reduction at Level Crossings, and analysed the combination of obstacle detection sensors with intelligent decisions layers, e.g., Deep Learning (DL) models, as an opportunity to provide robust interlocking decisions.

These solutions aim at increasing the safety level at LCs, however, in our view, guaranteeing the correct functioning of LCs should be the first step to ensure safety and traffic availability. Therefore, in the following of this chapter, we will focus on delineating how AI (specifically Deep Learning) can be integrated to support and improve current maintenance practices at Level Crossings to move towards proactive and predictive maintenance activities.

4.2. Background and Description

From a high-level perspective, Level Crossings can be subdivided into two macro categories: *Passive LCs*, which involve neither warning signals (warning bells or warning lights) nor barriers, and *Active LCs*, which can involve both warning signals and barriers. In addition, *Active LCs* can be *Automatic*, i.e., the barriers are automatically triggered by the approaching train, or *Manual*, i.e., the barriers are manually activated by human operators. In the following of the RAILS project, we focus on **Automatic Active LCs** involving all the three macro components mentioned above: warning bells (WBs), warning lights (WLs), and barriers (Bars). All these components must work properly together respecting particular constraints (e.g., [8]) to ensure safe operability. The specific requirements, e.g., the time the Bar should take to close and similar, will be analysed in the following phases of the project. In this document, we mainly focus on understanding how and through which sensors it would be possible to monitor LCs in near real-time.

[1] https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Railway_safety_statistics_in_the_EU

Typically, LC inspections are conducted on the field by following a fixed scheduling (see, for example, [9, 10]), however, failures may occur between adjacent inspections. This is one of the motivations that, in the last years, contributed to increasing the interest in continuous monitoring systems to collect data related to the health status of LC systems[2,3] in real-time to promptly identify malfunctions. The main issue, in this case, is related to the installation of intrusive IoT sensors (i.e., applied directly on the asset to be monitored); if, from the one hand, they represent one of the main enablers for real-time monitoring and the realisation of Digital Twins [11], on the other hand, they installation introduces two main complications when it comes to LCs:

- First, the newest LCs may already be equipped with the latest and adequate sensors, however, there are thousands of "old" LCs, deployed decades ago, which may be not; therefore, this would lead to a massive, expensive, and time-consuming sensorization.
- Second, being safety-critical assets, level crossings are subject to strict standards and regulations including, for example, CEN-CENELEC standards for the assessment of the Safety Integrity Level (SIL) [12]. In this context, the introduction of intrusive IoT sensors could lead to expensive and time-consuming re-approval processes.

IoT sensors could allow the real-time collection of data including, for example, the speed of the Bars or the current signals related to the activation of the WBs/WLs. Such data could (we will refer to as "primary data") be used, after a first simple processing step (e.g., cleaning, transformation), to build a Digital Twin of the LC and analyse its health status. However, given the issues presented above, alternatives would be required to extract about the same primary data leveraging non-intrusive sensors. To that aim, in our view, AI would be extremely useful. For example, assuming to have a camera monitoring the LC, AI approaches could be adopted to extract features related to the movement of the barrier from video frames. We will refer to these kinda of data (e.g., videos) as "secondary data", i.e., data that require an advanced pre-processing (e.g., AI based) to be used for the analysis of the health status of the system and/or the realisation of its Digital Twin. Hence, instead of using intrusive sensors (e.g., accelerometers) to directly collect primary data related to barriers' movement, it would be possible to combine secondary data from cameras and AI-based pre-processing to extract about the same movement data (this and other related aspects will be better discussed in the following of this chapter as well as in the next deliverables of the RAILS project). Therefore, practically, AI could enable the extraction of primary data from secondary data collected through non-intrusive sensors and, consequently, allow for the realisation of Digital Twins avoiding the aforementioned problems.

In this document, we intend to provide a contribution towards the continuous monitoring of LCs by means of cost-effective and non-intrusive sensors. Particularly, we will focus on camera and microphone sensors as: i) they are non-intrusive, i.e., they can be installed without compromising the functioning of the LCs; ii) they may be already installed at LCs for surveillance purposes, therefore, we could have some advantages related to technology reuse. Actually, the same applies in case they will be installed for maintenance purposes and leveraged, in the future, for safety-related aspects (e.g., obstacle detection). Eventually, data collected by means of these sensors could be processed with the help of AI to extract primary data that would allow for the realisation of LC Digital Twins.

[2] https://www.smartmotors.org/level-crossing-monitoring

[3] https://www.voestalpine.com/railway-systems/en/products/rxm-rail-crossing-monitoring/

On the bases of this analysis, we defined in Table 5.1 three main Key Performance Indicators (KPIs) that should be taken into account when evaluating LC continuous monitoring systems.

Table 4.1: KPIs for the evaluation of LC continuous monitoring systems.

KPI	Description
KPI1	**Implementability Costs**. Includes costs of sensors, installation, possible re-approval processes, and consumption.
KPI2	**Computation Time**. Indicates the time required by the system to detect possible malfunctions.
KPI3	**Effectiveness**. Indicates how the monitoring system would perform in terms of accuracy and detectable malfunctions.

There are two aspects to consider in relation to the above-mentioned KPIs. In particular, the continuous monitoring may be intended as *monitoring the LC when it is working*, this means that the system may be set on stand-by and activated only when the approaching train triggers the LC; in this way, we would also have savings in terms of power consumption (KPI1). Also, the real-time detection (KPI2) of possible malfunctions may not be a strict requirement; the purpose of these systems would be to detect defects to act before the system fails. This means that even though the system takes a few minutes to produce the output, it would be acceptable. Differently, if the intent would be to provide the incoming train with real-time indications for safety-related aspects (e.g., the Bar has failed, stop the train) KPI2 would assume more relevance. However, in this case, we are more in the context of safety than in the maintenance field, hence, other complications will arise related, for example, to the certification of the monitoring systems themselves against safety standards.

In the remainder of this chapter, we are intended to understand whether DL approaches can be exploited in combination with data from non-intrusive sensors (e.g., cameras and microphones) to identify the health status of any kind (old and new) of LCs from the maintenance perspective.

4.3. Objectives and Research Questions

Through this analysis, we intend to define some directions that go even beyond the context of the proposed case study. Indeed, most of the concepts and challenges that will be discussed in the next sections have a broad spectrum and characterise/affect different AI and railway applications. Therefore, the case study we intend to address should be considered as a means to contextualise the definition and (visionary) the resolution of some of the challenges that are currently limiting the adoption of AI in railways. At the same time, in the specific context of LCs, we intend to: i) define a methodology to schematise constraints, requirements, and directions; ii) identify necessary tools to overcome particular issues; iii) highlight already existing datasets (if any) and AI models that could be exploited to deal with LCs predictive maintenance.

To that aim, taking also into account Section 4.2, we defined three Research Questions (RQs) to structure the analysis:

RQ1: How can we automatically detect anomalies in LC devices, such as Barriers, Warning Lights and Warning Bells, for maintenance purposes, by using AI applied to possibly existing LC surveillance cameras and microphones?

RQ2: How can we generate relevant datasets for Warning Lights, Barrier movement, and/or Warning Bell anomaly detection based on Deep Learning?

RQ3: Can we demonstrate possible answers to RQ1 and RQ2 through a simple proof-of-concept demonstrator in order to inspire future developments and a technology roadmap?

4.4. Methodology and Tools

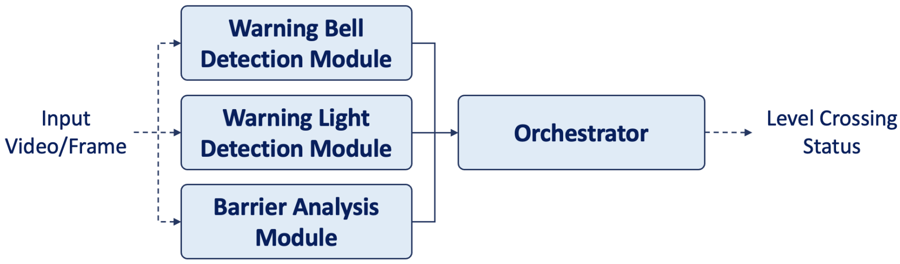

Fig. 4.1. High-Level Framework for Level Crossing Monitoring

As mentioned in the Sections above, the focus of this research task is to understand whether the health status of Automatic Active LCs can be determined by means of DL techniques and data coming from non-intrusive sensors such as cameras and microphones. As Automatic Active LCs are composed of three main components, in our view, it would be advisable to build a multi-modular framework that expects each module to monitor, in parallel, a specific LC component (Figure 4.1). In particular:

- The **Warning Bell Detection (WBD) Module** will be oriented at detecting the WB.
- The **Warning Light Detection (WLD) Module** will be oriented at detecting the WL.
- The **Barrier Analysis (BA) Module** will identify the motion of the barrier and detect whether it deviates from its correct behaviour.
- The **Orchestrator** will coordinate the three modules listed above in order to establish the health status of the whole LC system.

Worth highlighting, **this framework would have the great advantage of allowing us to consider the whole Level Crossing System as a black-box, focusing only on its behaviour, i.e. just evaluating its visible and audible output and estimating its deviations from the expected/nominal functioning.**

In order to realise a proof-of-concept based on this framework, some enabling tools would be required in case it is not possible to perform actions on the field, e.g., installing sensors at LCs to collect data. Particular concerns are related to the BA module since it is not possible to build a suitable dataset by leveraging online repositories (as we will see in the following section).

Therefore, research-oriented simulators or other tools like real-world simulators (i.e., software born as video games but exploitable to make ad-hoc simulations) that would allow for comprehensive data collection, e.g., under different environmental conditions or different operability conditions, would be required for this specific task.

For research-oriented simulators, we mean simulators specifically designed for scientific/testing purposes (e.g., Simscape[4] or RoadRunner[5]). These clearly offer advantages when it comes to simulating, i.e., meticulously defining, the behaviour of the systems. Indeed, taking Simscape as an example, it would be possible to specifically define the behaviour of the various objects (e.g., Bars); hence, it would be possible to simulate both normal and abnormal behaviour. However, these simulators may lack graphics characteristics, which is one of the main aspects to consider when it comes to computer vision approaches. On the other hand, real-world simulators like Grand Theft Auto V (GTA V)[6] - which has been already investigated for self-driving cars [13–15]) - may offer near-real graphics and the possibility to efficiently and easily customise the environment in terms of weather conditions and external agents. However, the characterisation of the behaviour of the various elements (e.g., the blink of WLs and the movement of Bars) is quite tricky; these elements have been coded to work perfectly, so we would not be able to collect data related to failures. However, as discussed in the following sections, we could adopt some expedients in order to realise our proof-of-concept.

4.5. Reference Datasets

From a data perspective, we can assume that each of the modules indicated in Figure 4.1 can be trained and tested with different datasets. Therefore, in this section, we discuss how these datasets can be built also considering that *no ad-hoc dataset seems to be available online [16]*. However, it would be possible to exploit online repositories or platforms to quickly build a dataset to train or pre-train (in combination with transfer learning approaches) models for proofs-of-concept.

4.5.1. Dataset for Warning Bell Detection

In the case of WBs, it would be possible to exploit already existing collections of data (e.g., AudioSet [17, 18]) and/or leverage recordings from online repositories/platforms (e.g., YouTube[7]) to build a suitable dataset.

AudioSet collects over two million audio samples (most of which have a duration of 10 seconds) and subdivides them into more than 600 classes[8]. As far as we know, it does not involve a specific class for LC WBs, however, it would be possible to retrieve data related to kinds of sounds that may be easily confused with LC WBs which would be extremely useful to improve the robustness of the AI system. Therefore, in our view, a good practice would be to build a multi-class classifier trained on a dataset composed of three classes:

- The **Warning Bells (WB)** class, which clearly encompasses the WB sounds.
- The **No Alarm (NA)** class, which encompasses audios which do not record any kind of alarm.
- The **Generic Alarm (GE)** class, which encompasses audios related to sounds that can be recorded near LCs and can be easily confused with WBs.

[4] https://it.mathworks.com/products/simscape.html
[5] https://www.mathworks.com/products/roadrunner.html
[6] https://www.rockstargames.com/it/gta-v
[7] https://www.youtube.com
[8] https://research.google.com/audioset/

The necessity of having three classes derives from robustness concerns. Typically, it is possible to approach these kinds of tasks by leveraging a binary classifier; i.e., a model capable of discerning acoustic activities related to WBs from acoustic activities related to other kinds of sounds (or no sounds). However, LCs are widespread in different environments, from cities to the countryside, where other sounds like sirens, alarms, or bells could interfere with the classification as they may be too similar to WBs. Hence, if we had only two classes (i.e., WB and NA), there would have been a high risk that most of these sounds would have been classified as WB. Therefore, the third class (GE) is almost necessary as it would allow us to mitigate this risk. Worth highlighting, the kinds of sounds that the GE class should encompass strongly depend on the specific LC WB under examination. For example, some British WBs could be confused with car alarms, ambulance/police) sirens, and other kinds of sirens/alarms; differently, some Italian WBs are more similar to bells (e.g., church bells, bicycle bells, and even cowbells). Therefore, GE samples must be selected ad-hoc.

Lastly, in our understanding, each country adopts a different sound for LC warning bells and, actually, we could also have different kinds of WB sounds within the same country. Therefore, building a comprehensive dataset encompassing, for example, all the level crossings deployed within the countries of the European Union would be extremely complicated and also counterproductive. Overloading a classifier with too many features to learn could bring to poor performance. Therefore, even if it sounds trivial, it would be advisable to build a different classifier for each country to reduce the aforementioned overload and the sub-classes of audios that the GE class should encompass.

4.5.2. Dataset for Warning Light Detection

To detect Warning Lights, in our view, it would be possible to proceed in two main ways: i) by adopting an object detector approach which will locate and classify the WL in “on” or “off”; or, ii) by adopting a simple CNN classifier. The second approach comes from the fact that our system is based on data coming from a fixed camera, which means that a WL occupies always the same position within the video frames. Therefore, it would be possible to crop the WL out from the image and then classify it as “on” or “off”.

The second approach would be easier to implement and, probably, faster in the inference phase. However, the choice would also depend on which kind of data it is possible to collect. As a general role, it would be advisable to collect data related to a specific LC and train (or fine-tune) the DL model on these data.

Despite the fact that, when it comes to AI, it is not easy to select the best approach a-priori; in this case, a classification approach would be preferable as it would be much easier to build an ad-hoc dataset. Indeed, to train an object detector (e.g., YOLOv5[9]), we would need a labelled dataset that, for each of the WLs, indicates the position within each frame (i.e., the bounding box) and the class (i.e., “on” or “off”). Therefore, it would be extremely time-consuming to build such a dataset from scratch. On the other hand, the dataset for the classification approach would be much easier to collect (even from scratch), as we would simply compose two collections of images: the first one containing switched-off WLs and the second one containing switched-on WLs.

Worth highlighting, when it comes to Traffic Lights (automotive), there are some datasets available online that could be leveraged to cope with the detection task.

[9] https://github.com/ultralytics/yolov5

For example, the LISA Traffic Light Dataset[10], the Bosch Small Traffic Lights Dataset[11], the SJTU Small Traffic Light Dataset[12], but also other well-known datasets such as COCO[13], can be exploited to pre-train the object detector model to subsequently apply a Transfer Learning approach. Notably, some of them also encompass bounding box annotations.

On the other hand, it seems that no ad-hoc dataset is available for a classification approach. However, it would be quite trivial to extract a suitable dataset starting from the video frames we would collect to address the Barrier Analysis task (discussed in section 4.5.3).

As for environmental characteristics the dataset should include, it depends on the meaning we would like to attribute to the LC Intelligent Monitoring System (Fig. 4.1):

- In case the system would be adopted for maintenance purposes only, it could be reasonable to think that it should not meet strict real-time constraints. At the same time, there would not be the necessity to constantly monitor the health status of the LC in each instant of the day. Therefore, the dataset could not encompass data related to poor light or visibility conditions (e.g., during the night-time or when it is foggy). Worth mentioning, the poor visibility due to night-time could be mitigated by adopting night vision cameras.
- On the other hand, in case this system would be adopted to also warn the train - in real-time - as to start automatic procedures to, for example, stop it in case of malfunctions (i.e., for safety-critical purposes), it would be required that the dataset contains samples related to all the possible weather and light conditions as to allow the model to properly take a decision in every scenario.

4.5.3. Datasets for Barrier Analysis

For this specific task, unfortunately, it would be extremely difficult to leverage online repositories. Even though different videos of LCs can be found on YouTube, there is a particular characteristic, as already explained above, that can help ML/DL algorithm to achieve better performances but is not fulfilled by videos available online.

Practically, similarly to what happens to WLs, when using a fixed camera, the portion of the frame occupied by the barrier is cyclically invariant. In other words, if the barrier occupies a given position when closed, once it opens and then closes again it will occupy the same position as before. The same applies to every possible position of the barrier. This means that the algorithm, whichever it is, will be somehow facilitated by the fact that it should “search” for the barrier only in a sub-portion of the image. Also, if we train the system with a dataset tailored to a particular barrier, the position of the barrier will be the same both in the training and inference phase; which simplifies further the task for the ML/DL model. To that aim, there are two possible solutions: i) acquire data on the field by using a fixed camera, or ii) exploit research-oriented or real-world simulators by keeping the visual fixed.

In our case, we will most likely take into account the second alternative. Up to now, we have started testing an object detection approach by leveraging data collected through GTA V (which bounding boxes have been manually added to train the network), however, we have encountered some difficulties in simulating the anomaly behaviour of the barrier. Hence, most likely, we will investigate the possibility of artificially creating anomaly data (e.g., by

[10] https://www.kaggle.com/datasets/mbornoe/lisa-traffic-light-dataset
[11] https://hci.iwr.uni-heidelberg.de/content/bosch-small-traffic-lights-dataset
[12] https://github.com/Thinklab-SJTU/S2TLD
[13] https://cocodataset.org/#explore

adequately shuffling video frames obtained from GTA V) or leveraging other kinds of simulations. As for data diversity in relation to weather and light conditions, the same considerations made in Section 4.5.2 hold also in this case.

4.5.4. Summary of Datasets, Repositories, and Simulators

Basing on what has been discussed so far, Table 4.2 summarises and briefly describes the datasets and simulators that could be leveraged to address the implementation of the modules indicated in Fig. 4.1.

Table 4.2: Summary of Datasets, Repositories, and Simulators

Module	Dataset / Repositories	Simulators
Warning Bell Detection	No dataset directly exploitable has been found. Repositories/platforms that can be leveraged to build a suitable dataset are: AudioSet [17] and YouTube[a].	In our view, it would be better to exploit real-world data (with related background noises) instead of synthetic data collected through simulators.
Warning Light Detection	Potentially, any automotive dataset containing traffic lights can be exploited to extract suitable data for this task. Possible candidate are: the LISA Traffic Light Dataset[b], the Bosch Small Traffic Lights Dataset[c], the SJTU Small Traffic Light Dataset[d], but also other well-known datasets such as COCO[e],	No need for simulators. Available datasets and repositories seems to be enough to build ad-ho datasets to properly train and test AI applications.
Barrier Analysis	No dataset directly exploitable has been found. Repositories as YouTube[a] could allow for real-world data collection, but they usually lack some task-specific characteristics (e.g., a fixed camera) that would help to facilitate the detection task.	Research-oriented simulators (see Section 4.4) may lack adequate graphic characteristics (key factor when applying computer vision approaches). Real-world simulators (e.g., GTA V[f]) represent a suitable solution to properly collect data in this case.

a. www.youtube.com
b. https://www.kaggle.com/datasets/mbornoe/lisa-traffic-light-dataset
c. https://hci.iwr.uni-heidelberg.de/content/bosch-small-traffic-lights-dataset
d. https://github.com/Thinklab-SJTU/S2TLD
e. https://cocodataset.org/#explore
f. https://www.rockstargames.com/it/gta-v

4.6. Artificial Intelligence and Machine Learning Models

In Section 4.4, we have proposed a multi-modular architecture to estimate the LC health status. Excluding the “Orchestrator” module, which will implement the logic to gather the output of the other modules and compute the general status of the LC, for each of the other modules it is reasonable to adopt DL approaches. The main reason lies in the fact that DL approaches (e.g., CNNs and Object Detectors) have shown great results compared to traditional ML approaches or other AI techniques when it comes to analysing audio-video data for maintenance purposes [19].

By leveraging findings and results obtained within RAILS’s WP1 encompassing a generic literature review of AI in railways [20, 21], an ad-hoc literature review on audio-video based defect detection through DL in railway maintenance [19], and the transferability analysis performed in the first step of RAILS’s WP3 [2], we propose here a brief discussion oriented at identifying which, in our view, would be the “best practices” and most suitable AI approaches to implement the modules depicted in Fig. 4.1:

- **Warning Bell Detection**. As discussed in Section 4.5.1, it would be possible to exploit the AudioSet [17, 18] dataset and YouTube to build a suitable dataset for WB detection. In addition, it would be possible to leverage a VGG-like architecture (named VGGish [17]) developed in the context of the AudioSet dataset, to build a CNN-based classifier to address our task. In brief, the CNN will take in input a portion of the audio and classify it into WB, NA, or GE as already introduced in Section 4.5.1. Furthermore, it is also worth highlighting that VGGish is provided with some pre-trained weights[14] which would allow us to also investigate the potential of Transfer Learning [22] in this context.

- **Warning Light Detection**. As indicated in Section 4.5.2, this case study has a peculiarity that other applications may not have: the camera monitoring the level crossing is fixed. At the same time, also warning lights are fixed in position. Therefore, we could proceed by (i) adopting an approach based on object detectors (e.g., R-CNN [23]) or (ii) adopting a simple CNN-based classifier. Indeed, given that the WLs are fixed in position, it would be possible to crop them out from the original image in a pre-processing phase and then leverage the classifier (which may also be a trivial Artificial Neural Network) to recognise the status (i.e., "on" or "off"). Identifying the best approach a-priori is not straightforward. On the one hand, a classification approach would suffer a possible shift of the camera: if the field of view of the camera changes due to external factors, the pre-processing step (in particular, the cropping) should be re-calibrated. On the other hand, an object detection approach would overcome this problem, but it may suffer complex backgrounds or the fact that WLs typically occupy a small portion of the image (small-scale object detection problem [19]). Possibly, a classification approach would be advisable in this case given the reduced network complexity and the fact that, in our opinion, the "shift" problem would be easier to tackle (by simply re-calibrating the cropping area) than the small-scale object detection problem.
- **Barrier Analysis**. As for WLs, also in this case we could crop out from the whole frame only the sub-area within which the barrier moves. However, for this kind of "motion-based" tasks, as defined in D3.1 [2], it seems that limited or no work has been carried out in the literature [2,21]. In our view, a suitable approach would be to adopt an object detector (e.g., YOLOv5) to detect the barrier within adjacent frames and then extract features related to the movement of the barrier by taking into account, for example, the height of the different bounding boxes. Then, we compare the nominal path the barrier should trace with the path in output to the object detector. This approach is better discussed below.

Fig. 4.2 expands the architecture we presented in Fig. 4.1 and depicts what is discussed above. By zooming into the Barrier Analysis module, we would probably study an approach as represented in Fig. 4.3. Practically, the object detector is used to extract the bounding boxes of the barrier over different frames; then, we plot the heights of these bounding boxes to obtain a representation of the movement of the barrier; lastly, we compare the nominal movement of the barrier with the one obtained leveraging the object detector. Hence, a Bar health score is presented in output.

Which specific technique should be used to compare the two paths is still to be defined and would represent one of the outcomes of the following RAILS tasks. This is because it would be possible to leverage traditional ML approaches, recurrent neural networks, but also approaches that do not leverage AI (e.g., simply computing the mean squared error - MSE - between the two paths). Hence, results-guided analysis should be carried out to understand which would be the best practices when it comes to dealing with motion-based health status estimation.

[14] https://research.google.com/audioset/download.html

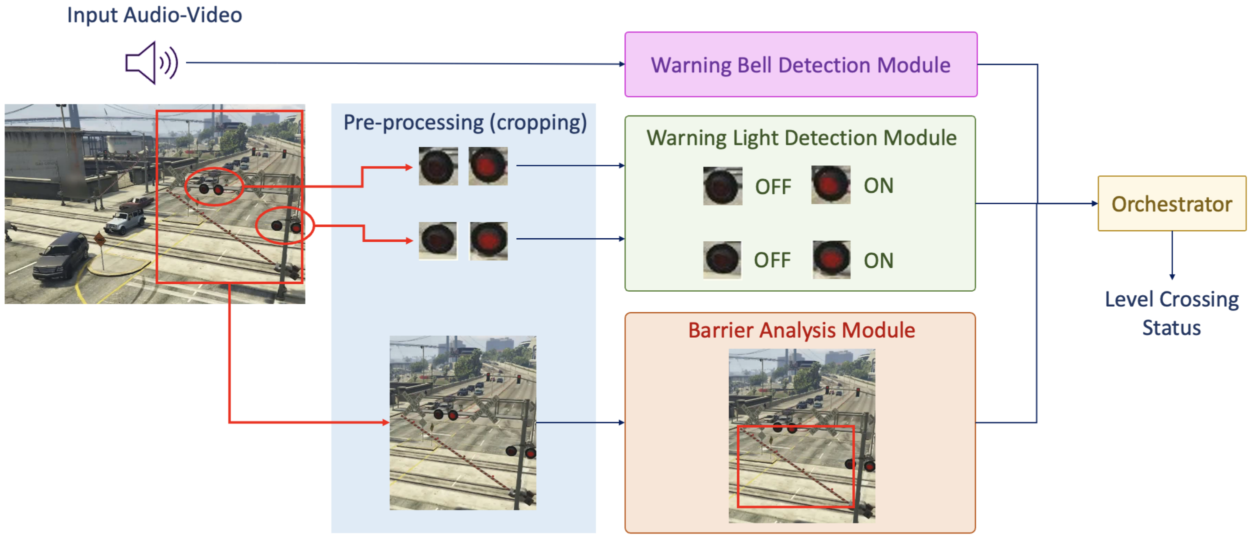

Fig. 4.2. Level Crossing Intelligent Monitoring System

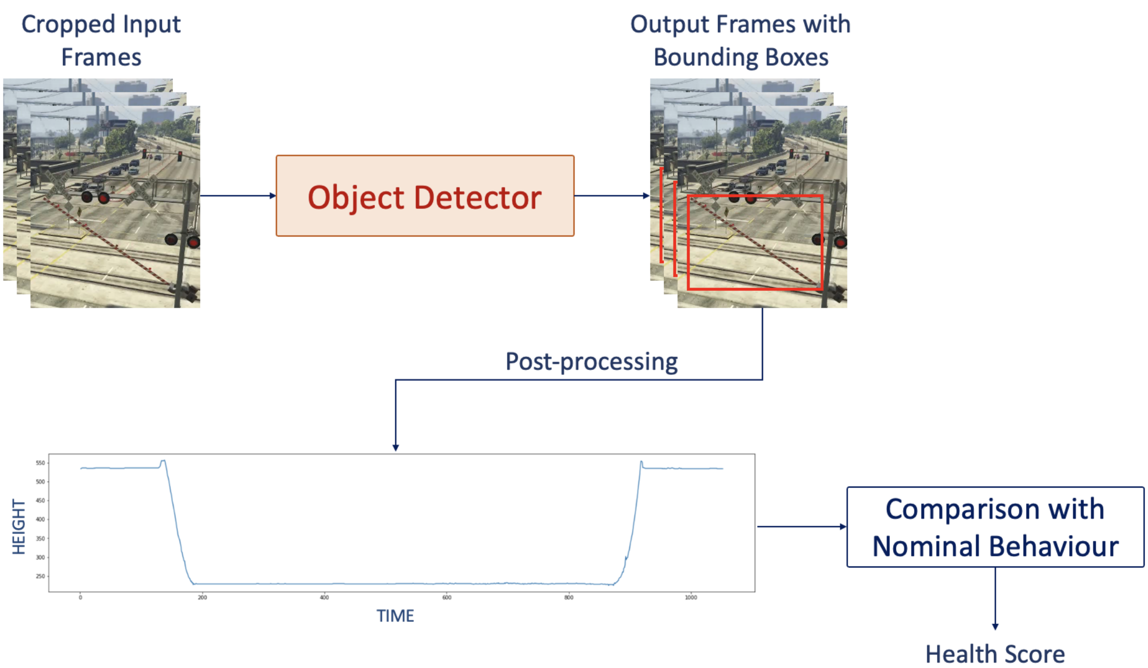

Fig. 4.3. Possible Framework for the Barrier Analysis Module.

Worth highlighting, the video frames represent what we called “secondary data” in Section 4.2, while the heights of the bounding boxes plotted over time (see the chart in Fig. 4.3) represent what we called “primary data”. Basically, with the help of a DL object detection model it would be possible to extract relevant data that can be potentially used, in addition to the prediction/detection of possible malfunctions, also for the realisation of a Digital Twin (for the barrier in this particular case). Indeed, while an intrusive IoT sensor (i.e., an accelerometer) could give us information about the “speed” of the barrier, with the approach hypothesised above we could retrieve information about the position of the external edge of the barrier; in both cases, we would be able to achieve about the same results, i.e., the collection of motion data in real time that could help us to both let the Digital Twin evolve with its physical counterpart and detect potential anomalies within the movements.

To conclude, Table 4.3 summarises what was written above and highlights the most promising classes of approaches we identified; column "contributions" reports a qualitative score (i.e., limited, medium, or high) indicating the coverage that, in our opinion and given the above and past [2, 21] analyses, the related topic has obtained within the literature or other projects.

Table 4.3: Suitable AI approaches

Module	Contributions	Suitable Approaches
Warning Bell Detection	Medium	CNN-based classifiers
Warning Light Detection	High	CNN-based classifiers (advisable in combination with a fixed camera) or Object Detectors.
Barrier Analysis	Limited	Object Detectors to extract motion-related features and another technique (to be defined) to extract a health score.

4.7. Expected Results and Possible Criticalities

This case study helps us to investigate methodologies and AI approaches that could be leveraged to deal with non-intrusive maintenance tasks. As briefly mentioned, in case the system will reach an adequate level of accuracy, trustworthiness, and reliability, it would also be applied for safety purposes, i.e., implement automatic safety functionalities to stop (or slow down) the train approaching to the LC in case of malfunctions. Another advantage is related to the sensors we are considering, i.e., cameras and microphones. Indeed, it would be possible to exploit these sensors (mainly cameras) also for other safety-related aspects, e.g., intrusion detection within the LC area to avoid possible collisions between trains and road users. In addition, the non-intrusiveness and affordability of these sensors, compared to others, may also allow for the easy deployment of this system without running into possible re-approval processes.

From a methodological perspective, this case study would allow us to evaluate Transfer Learning approaches and investigate approaches to deal with motion-based health status estimation of moving objects which, as already stated within D3.1 [2], seems to be still an open issue in railways as well as in other sectors.

The first step we intend to take in the future is to identify exploitable datasets, among those cited in this chapter, that could be useful for our purposes. Then, by leveraging these datasets or additional sources, we will build ad-hoc datasets to properly train the discussed modules (i.e., WBD, WLD, and BA). Hence, as the main future outcome, we expect to delineate roadmaps and guidelines for the health status estimation of railway assets starting from the realisation of the proposed proof-of-concept, i.e., the realisation of the proposed modules. For example, a reasonable question we would like to answer would be: how can the health status of a moving object be estimated leveraging video data only?

To conclude, the approach we propose encompasses some practical criticalities that we will most likely face, and possibly address, in the next phases of the project. First, once the WB or the WL has been detected, to estimate the functioning of the physical system (i.e., the semaphore or the bell) we should also analyse temporal dependencies. For example, the

WL should blink at intervals of X milliseconds: how can we estimate if it blinks "correctly" starting from the WL detection? Perhaps, it would be possible to leverage an approach similar to that presented for the Bars, i.e., turning the detection into temporal data and then comparing these with the nominal behaviour. However, this represents the second criticality: how can the detected behaviour be effectively "compared" with the nominal one? Are ML or DL approaches required or it is possible to estimate the health score through non-AI-based techniques? Lastly, as a more visionary outcome: would it be possible to estimate the Remaining Useful Life (RUL) of these systems besides the estimation of their health status based on qualitative scores (e.g., healthy, unhealthy, failed)?

4.7.1. Opportunities for AI-enabled Digital Twins and Self-Protection Mechanisms

Despite not being the subject of our future research, it is important to underline the potential of AI combined with Digital Twins (DTs) when it comes to smart maintenance applications.

As [11] outlines, AI could act as one of the main enablers when building DTs and it can be leveraged in the different architectural layers composing the DT [24]. Without going deeper with the details, in our view, AI could play two main roles when building a DT; it can be involved to i) pre-process what we defined as secondary data (data extraction); and ii) build behavioural models of physical assets that can be used to make predictions (modelling/data analysis). Thus, in this context, we expect that the algorithms and approaches we will develop/adopt could be potentially useful in both these cases. For example, as already stated in the sections above, the barrier detection approach (in the context of the Barrier Analysis module) can be seen as an AI-based pre-processing step. On the other hand, the WBD module would implement something that is more similar to a behavioural model of the Warning Bell(s) to detect whether there is sound activity or not.

In addition to that, LC DTs would open for new opportunities at the rail network level. Assuming that, somehow, it is possible to obtain a digitalized representation - in terms of DTs or simple data streams - of all the entities composing a railway network (including LCs), it would be visionary possible to introduce some self-protection and/or self-adaption mechanisms as to improve, besides maintenance activities, also the *safety* of Level Crossings. Such a digitalization can be represented in terms of **levels of intelligence** as in Fig. 4.4 (excerpted from [25, 26]). We would have: i) *edge intelligence components*, oriented at actuating decisions locally to the asset they are deployed on (e.g., a train collects some environmental data and directly takes autonomous decisions about the most optimal action to perform); ii) *fog intelligence components*, which basically manage and control a specific railway track by taking into account data coming from the lower level and communications from trackside equipment (this is, most likely, the level at which LC DTs would operate); and iii) *cloud intelligence components*, which gather data from multiple sources (possibly deployed worldwide) to build larger knowledge bases on which rely for smarter predictions and maintenance interventions.

In this context, in case of malfunctions of the LC or, in alternative, after a deduction made by stressing the LC DT (e.g., after testing potential outcomes through fault injection [27]), it would be possible to (manually or, visionary, automatically) activate some (self-)protection mechanisms as depicted in Fig. 4.5. Practically, by exploiting the DT, it would be possible to understand which actions should be taken at the railway network level to avoid unpleasant consequences. So, if the LC DT detects a critical malfunction (e.g., the barrier does not move), we could potentially have a *Track Controller*, that is connected to all the DTs of that

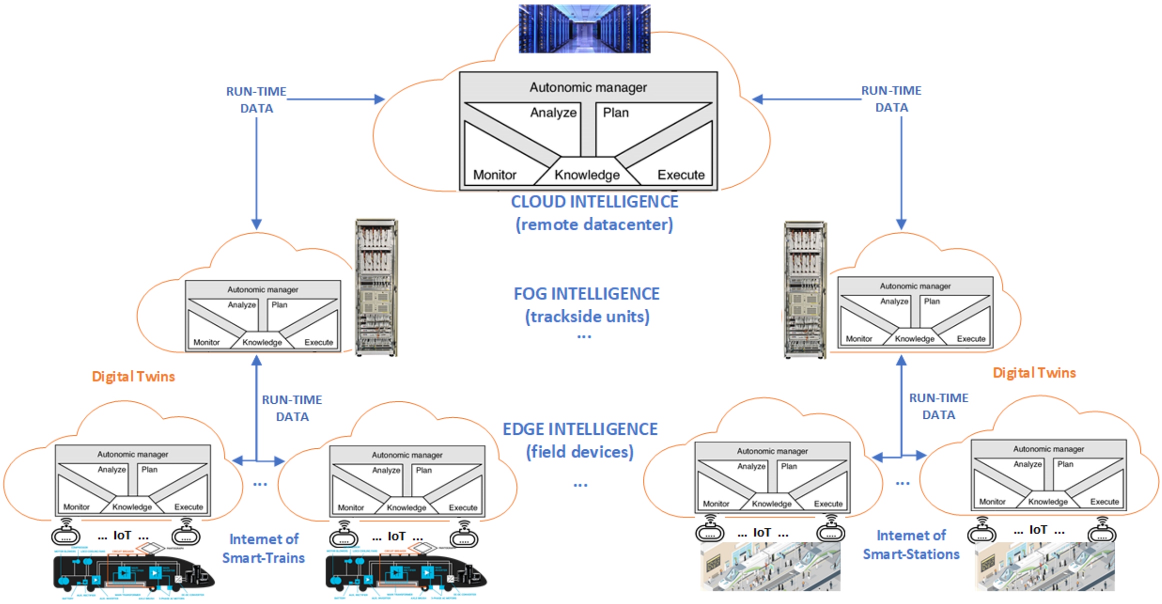

Fig. 4.4. Levels of Intelligence for Digitalized Railways [25, 26].

specific rail track or may be itself a DT of the whole track, which receives the notification from the LC DT and takes the most suitable decision basing on data coming from other DTs (e.g., the one of the train in the figure) and the upper layer. Hence, at this point, we could mainly have two alternatives. The first possibility (as shown in the figure) would be that the *Track Controller* directly communicates the action to be taken to the device controllers (the *Train Controller* and the *LC Controller* in this case), for example: the train should proceed at reduced speed or should stop, while the LC should activate some warning alarms to the road users. Alternatively, the *Track Controller* could communicate the actions to the DTs which, in turn, will propagate the actions to the device controllers.

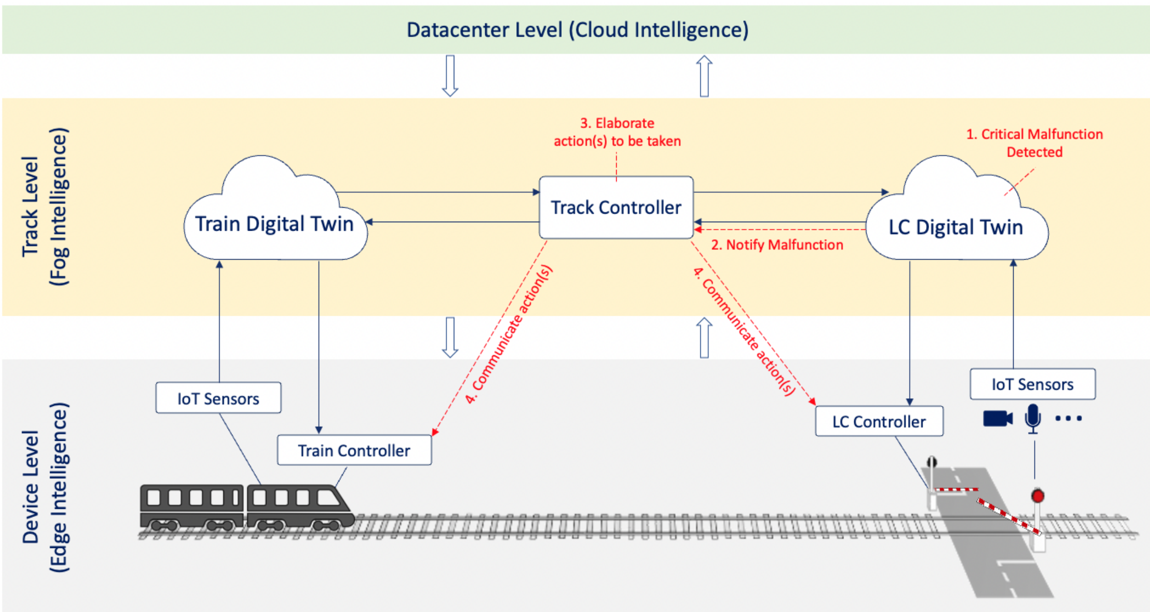

Fig. 4.5. Self-Protection Mechanism in case of LC Failure.

To conclude, what discussed above is just an example on how DTs and levels of intelligence could be leveraged to monitor and control the entities (trains, LCs, etc.) operating on a specific railway line in the proximity of a defective LC. Fuhrer analysis would obviously be required to specifically define the architecture of the DTs and how the communication should be made. However, in the following of the RAILS project, we will mainly focus on the investigation and implementation of the AI approaches (discussed in Section 4.6) that could visionary became enablers for the realisation of Digital Twins, rather then on the implementation of Digital Twins themselves.

5. AI-based Predictive Maintenance for Rolling Stock

5.1. Introduction

In the very last part of deliverable D3.1 [2], we proposed two potential pilot case studies that may be worth to explore further. One of it - Smart maintenance for Level Crossings - already been properly discussed and in chapter 4. This chapter will particularly focus on another: Predicting the health condition of rolling stock components, especially wheels and axle bearings, based on the trajectory of vehicle. Which is inspired by the high-level idea we already proposed in D3.1 of exploring digital twins applications being implemented on railway station infrastructure/system/component for maintenance purpose. From the high-level perspective, rolling stock maintenance is a significant aspect of the maintenance activities occur in railway stations. The current maintenance type performed on rolling stock components are typically intrusive maintenance, and such maintenance has to be conducted in the depots or station hubs, on a static manner. With these considerations, we narrow down our scope into the AI-based Predictive Maintenance for rolling stock components, by borrowing ideas we found in digital twins.

Railway axle bearing defects can reduce operational efficiency or cause in-service breakdowns, causing track and train damage. The majority of railway accidents are caused by wheel and bearing failures, which result in train derailment. Healthy bearings generate a particular degree of vibration and noise, however a defective bearing produces significant changes in vibration and noise levels. Bearing faults can be detected early in their development, allowing an operator to rectify the damage before it becomes severe. When a vehicle is due for maintenance or overhaul, knowing the extent of bearing deterioration is advantageous, as it leads to fewer operating issues and increased fleet availability.

Significant amount of research has been carried out in recent years to improve railway fleet safety. From the perspective of data sources, most of the existing studies has been focused on the creation of sensors to monitor and collect the health condition of railway equipment (e.g. [28] and [29]). In terms of the solutions to the rostering and scheduling for rolling stock maintenance activities, multiple approaches have been investigated and implemented. For example, [30] created a genetic algorithm for the optimisation of rolling stock maintenance overhaul scheduling with more maintenance planning considerations. A more intelligent model of the train repair routing problem was presented by [31], where similar constraints on maintenance were included, however, solve the train rostering issue instead. Additionally, they point out that only long-distance trains should be the subject of a joint study of rostering and rolling stock maintenance. The maintenance routing problem was given a mixed integer formulation by [32], with the shunting process acting as the bottleneck. The scheduling issue for preventative maintenance and reducing the time needed to complete maintenance tasks were covered by [33]. By integrating the maintenance tasks performed on each track, heuristic algorithms construct almost optimal solutions, for instance, A multiple criterion decision-making dilemma was put up by [34], who also examined several equipment maintenance tactics. A branch and bound algorithm is used to solve the problem, which is presented as an integer programming process. However, from the above mentioned previous methods we can know that managing rolling stock equipment in railway efficiently and effectively is challenging due to it is a comprehensive task which requires a relatively high-reliable performance and the applied technologies need to be specialised for achieving the particular needs of any industrial and business goals. The emerging digital technologies and Artificial Intelligence are expected to argument the decision making in vehicle assest management and maintenance scheduling. In this deliverable we aim to borrow some concepts from "digital twins", into the tasks of helping transform how staff/maintenance team approach and interact with the rail systems/components.

A 'digital twin' application is a virtually-existed, data-driven representation of the all essential characteristics and behaviours of any real-world physical asset, operation process, or a system [35]. In the railway sector, existing promising attempts have been found across several different aspects including smart maintenance, planning management, and operational monitoring. Particularly, due to digital twins techniques spans the asset's entire life-cycle, one of the major areas of potential is rail infrastructure and rolling stock maintenance, especially with regard to condition-based maintenance. In fact, the use of digital twins in asset simulation can be very helpful for the prediction and identification of components in tracks and vehicles that pose the risk of failure. This ensures that the lessons learned thanks to these digital solutions contribute to design enhancement in line with operational requirements, and that they are applied to future products and systems.

Furthermore, during the planning and construction of a new railway or rolling stock – as well as upgrade or renewal processes – digital twins can contribute to the prediction of changes occurring in the project execution, as well as the risks brought by non-conformant construction. Therefore, using digital twins can help to drive maximum value from infrastructure and rolling stock investments, avoiding cost increases and delays. Furthermore, a digital twins ecosystem can be much broader than the core rail-related systems and subsystems. In fact, if we consider large rail infrastructure projects, such as 'Le Grand Paris Express'[1], these ecosystems can be extended towards construction and civil work industries, as well as infrastructure planners and real estate managers.

It is clearly convinced that Digital Twins are helping transform how staff approach and interact with the rail systems. By developing a digital representation of a physical thing or the prediction procedure, they are capable of managing and analysing the ever-growing information available to us. In return, such virtual replica enables us develop a more comprehensive understanding of how is the degradation status of railway components, which helps us to clearly see how we can make maintenance activities more efficient [36].

With the growing popularity of data-driven railways and the use of measurement technologies to aid decision-making in the railway sector, the practicality and quality of such measurements are becoming more important. Traditionally, safety-related decisions in the railway industry have been decided using a combination of high-precision inspection against specified criteria and, when inspection is not possible, periodic action. One effective strategy for avoiding the wheel wear defects cause severe consequences throughout the world is to perform preventive maintenance, which involves opening machinery at a predetermined

[1] https://www.societedugrandparis.fr/info/grand-paris-express-largest-transport-project-europe-1061

interval regardless of component condition. Non-destructive testing (NDT) inspections are commonly utilised, however they are expensive, inefficient, and time consuming. Sometimes, it's possible that the maintenance team fail to notice the flaw, resulting in catastrophic failures [37]. Another obvious disadvantage is that the train must remain motionless during the inspection. While safety remains the top priority, maintenance decisions are now giving more weight to the bottom line. More and more organizations have use cases for digital twins combined with augmented reality (AR) only increases that justification. The benefits of such development during testing and validation are obvious. For example, AR can show a digital twin on top of a physical machine and provide information a technician wouldn't otherwise see, and technologists can enter the AR of a digital twin to simulate various issues. Continuous condition-based monitoring is utilised to guide decision-making in this field. Condition-based maintenance (CBM), which provides real-time and in-service measurements of railway components, is now the recommended method. Faults can be discovered while on the move thanks to this technology. This approach speeds up maintenance and expands the number of wagons/coaches available for use.

5.2. Background and Description

Generally, the applied industrial digital twins can be implemented under several scenarios: i) Monitoring physical assets, which includes the status and healthy condition of rolling stock, tracks and other infrastructure. ii) Monitoring train movements. For instance, developers are able to visualise railway fleet operations using data from OpenRailwayMap [2] and on-board sensors. iii) Providing information about passenger behaviour in vehicles, stations or on platforms. In this deliverable we tend to specifically focus on implementing digital twins-based prediction procedure into railway asset status monitoring especially the estimation of failure occurrence on wheels set and axle bearings. From this perspective, digital twins can help companies to recognize equipment failures before they stall production, allowing repairs to be made early or at less cost. Companies can then save even more money when they further automate their business response to such changing conditions.

Implementing AR/VR technologies in the digital–twin design and management process, however, can not only improve productivity within many industries — most notably aerospace, automotive, and industrial — it can also enhance experimentation and predictive analysis of existing products, with the potential of reducing downtime and incurred costs. Many factors influence whether wheel sets require maintenance, including hollow profiles, thin flanges, flats, rolling contact fatigue, and, in some cases, cavities or out-of-roundness. In general, Bearing defects can be classified into three categories [38]:

- Abrupt: symptoms appear unexpectedly
- Intermittent: symptoms are not continuously present; they can appear for a period of time and then disappear, which makes them challenging to identify
- Incipient: a fault which develops gradually over a period of time, with worsening symptoms

A vibration sensor installed on the axlebox housing or built into the axlebox bearing is used to monitor the condition of the wheel set by providing data on things such as wheel flats and wheel shape that may be used to evaluate the health status of the wheel set.

[2] https://www.openrailwaymap.org/

The shaft speed is a factor in the real-time calculation as well. It is incredibly expensive and time-consuming to maintain wheels. By maximizing the operational mileage of wheel sets, bogie condition-based maintenance enables the scheduling of these operations to be planned without sacrificing dependability and safety.

For decades axle-box bearings have been monitored in the railway industry by utilising stationary track-side mounted temperature/noise detection systems. This equipment is typically installed at certain intervals along the track or at strategic locations, such as ramps in the case of alpine railway lines. Such a system typically provides an indication of heavily worn or damaged components. As a result, the train must be stopped and the faulty wagon replaced and sent to the next suitable workshop. This causes operational delays and additional costs.

In 2014 a report by bearing manufacture Swedish Ball Bearing Factory (SKF) showed that causes of bearing damage can generally be divided into four main categories, as shown in Figure 5.1 [39]:

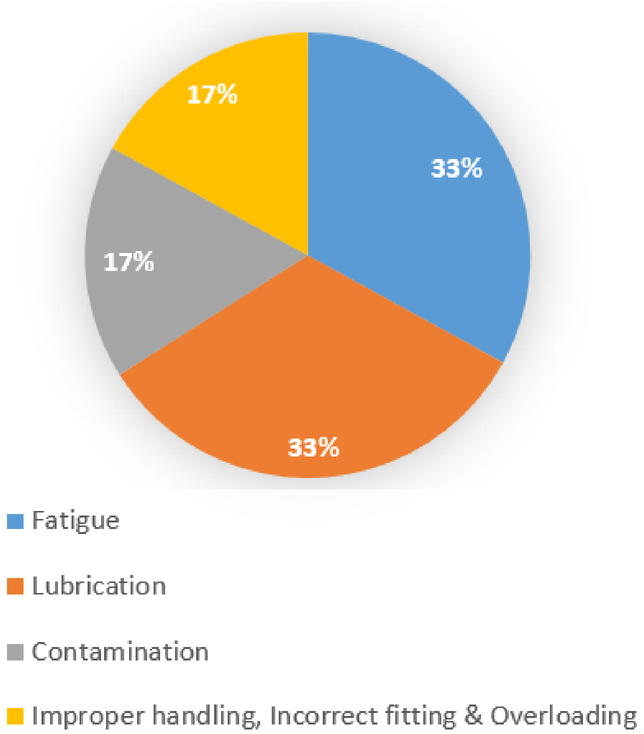

Fig. 5.1. Causes of damaged/failed bearings

Furthermore, most bearing manufacturers impose a mileage restriction between overhauls. The most common way of inspecting wheels is to measure profiles in depots during maintenance periods; this is a time-consuming and inexact process. Alternatively, automated systems have been greatly development during the past two decades. Bearings were assessed and replaced on a regular basis by the more accurate calculation of Remaining Useful Life (RUL), and hot axlebox detectors (e.g., [40]) or the image analysis techniques (e.g., [41]) were used to identify failures on rare occasions (HABD). The four stages of bearing wear have been defined according to [42]:

Stage 1: The wheel/bearing is damaged on a microscopic level, but it is still functional, and the chance of catastrophic failure is low. This level of damage is undetectable by most condition monitoring approaches.

Stage 2: The damage is obvious and some condition monitoring tools can detect it. At this point, it is advised that the repair be done during the next inspection. Failure is a possibility, but not a high one.

Stage 3: The damage is now evenly distributed among the rollers and races. A repair may be required within a month, depending on mileage (operational duty) and load. An out-of-service failure is a distinct possibility.

Stage 4: The damage is audible (to a nearby human). Immediate attention is required.

Because a heated wheel bearing is only hours away from breaking, a train with a suspected bearing is almost usually pulled out of service right away. The maintenance approach is presently changing from preventive maintenance to predictive/condition-based maintenance in order to reduce unplanned failures, operating and maintenance costs, and boost dependability and availability of rail vehicles. From this point of view, vibration monitoring, which can provide months of warning before failure, is a far superior strategy.

For implementing different sensors to diagnosis rail car component failures, both wayside and onboard sensor technologies is available and heavily investigated. For assessing axle bearing failure, wheel flatness, and wheel profile, many types of sensors such as temperature, vibration, acoustic, laser, ultrasonic, and force are employed either on-board or off-board. From the tactical perspective, the major discussion of contention appears to be whether track-side or vehicle-mounted monitoring is preferable. Both offer information on the condition of wheel bearings and wheels, but vehicle-mounted sensors can also provide information about the track and should be able to identify a hot axlebox, though with a significantly faster response time. Then there's the matter of how many track-side systems are required to cover an operator's whole network – a single track-side installation would almost certainly be insufficient for a large train operator with a vast network. For instance, [43] and [44] are two examples that using track-side acoustic emission detectors and wheel profile sensors to measure record and predict the performance of the specific components.

Through the previous investigation another observation inspired us to perform this case study: Rolling stock of vehicles would show a general pattern in their process of ageing and failures. Especially after the locomotive travelling through a 'notorious' track section or a certain area (i.e., Rough mountainous or overly steep areas), the probability of the occurrence of weak wheel contact and bearing overheating will greatly increase. That is to say, trajectory of the vehicle will probably interfere the performance of a specific component. Inspired by this, can we develop a digital twins simulator that incorporating such trajectory information into failure prediction and providing insights for predictive maintenance? We are investigating various analytical approaches, including correlation analysis, causal analysis, time series analysis, and machine learning techniques to automatically learn rules and construct failure prediction models using enormous volumes of historical detector data combined with failure data, maintenance action data, inspection schedule data, train type data, and weather data. To forecast future failure-causing circumstances and prevent service outages while enhancing network velocity, these models will be applied to both historical and real-time data. Although they do not directly address network velocity, the analytics and models can also be utilized by managers to proactively identify the root cause of various failure modes and component wear rate.

The three primary Key Performance Indicators (KPIs) that should be considered while evaluating the AI-based rolling stock predictive maintenance system are specified in Table 5.1 based on what we already introduced and in accordance with the prior case study.

Table 5.1: KPIs for the evaluation of rolling stock predictive maintenance systems.

KPI	Description
KPI1	**Implementability Costs**. Includes costs of sensors installation, equipment maintenance fee, and costs of collecting and storing necessary data.
KPI2	**Intelligent Degree**. To what extent does the implementation automate the process of analyzing and predicting the optimal maintenance scheduling is another significant indicator.
KPI3	**Effectiveness & Transferibility**. Indicates how the predictive maintenance system would perform in terms of prediction accuracy and how effective it can be normalized to another scenario.

There are three performance indicators have been considered in accordance with the above-mentioned maintenance frameworks. In particular, (KPI1) requires a comprehensive evaluation of the overall costs compared to conventional maintenance activities, which includes the consideration of whether/the most financially-saving locations to install the distributed accessible sensors, and how to make sure these supportive equipment are functional to properly collect and store data assets. Also, the automation degree (KPI2) of the proposed solution may be a significant aspect to consider. Despite the unexpected failure or systematic disruption, how likely is this system to correctly generate reliable maintenance decisions without the interventions of humans? Last but not the least, the expected outputs of this system are difficult to be measured quantitively however its effectiveness is concerned. People have interest in to what extent this improvement has brought to the cost-saving scheme, and how the innovative solution can be successfully transferred to another case study.

5.3. Objectives and Research Questions

A number of practices on predictive maintenance especially condition-based maintenance have recently been implemented, and their goal is to identify and suggest creative solutions using the intelligent distributed maintenance system, which supports the process of systematic preservation operations. To be more precise, discovering unexpected behaviours from vast amounts of equipment sensor data and turning them into machine-understandable and useful insights for the following proactive asset protection activities, such as averting potential accidents in advance via predictive perception. That is the fundamental tenet of predictive maintenance or intelligence forecasting, which places a strong emphasis on planning maintenance and repair procedures in advance of the loss of functional equipment.

In this deliverable, for automatic processing, diagnosis and discerning ageing, failures and faults on rolling stock components especially the wheels, axle bearings and chassis failure, a digital twins-based solution is going to be explored in our scope. In addition, predicting future failure modes so maintenance can be preemptively scheduled, and executed on other predictive maintenance goals, is the primary achievable goal of this case study. In this context, the tools and methods for supporting the adoption of a condition-based maintenance strategy for the chassis components of train locomotives—specifically, the wheelset and axle bearing—are being evaluated. This report, which includes findings from related projects, was created to provide an overview of the computational models and simulation tools that are now available for the study of the deterioration of railway bogie components. For example, the contact mechanics are also taken into account in a railroad vehicle simulation [45]. For the investigation of the wheel/rail phenomenon conditions, such as creep pressures, stress distribution, rolling contact fatigue, wear, analysis of the contact mechanics is utilized ([46] and [47]). Without investing the time, money, and resources necessary to test the railway vehicles on a track, a user can try out numerous scenarios using computer modelling ([48] and [49]). Modeling can anticipate the circumstances of a derailment situation, such as the speed at which a wheel or rail can derail or the circumstances in which it can be stopped ([49], [50] and [51]).

The decision of which computational models should be created and used in D3.3 will be supported by this review. Computer hardware and software have advanced significantly over the past forty years and are now utilised to simulate dynamic relationships in the field of railroad vehicle modelling. Universities, railroad organisations, and train manufacturers can use these simulation programs to help with their own projects and related issues, such as fewer (expensive) field measurements, parameter studies on the dynamic behaviour of the railway vehicle, and predicting and minimising the time and expense of improving the performance of the railway vehicle, such as increasing axle load and speed. It is possible to perform accurate models of the dynamic behaviours of railroad vehicles using contemporary computer software. A variety of computer programs have been created to assist engineers in modelling the dynamic characteristics of railway vehicles. These multiple tools were developed to offer environments that are easier to use, with additional graphical features, animations, and straightforward code. Almost all railway simulation software follow the same process. In the realm of railway vehicle component simulations, Simpack, VI-Rail (based on Adams/Rail), Universal Mechanism, as well as Vampire, Nucars, and Gensys, are now the most popular programs ([52], [53] and [54]). To forecast the normal and pathological behaviour of the vehicle and its components, these models will be utilised to execute dynamic simulations of the interaction between the vehicle and the track under actual operating conditions.

According to the considerations above, we proposed several Research Questions (RQs) for exploring further:

RQ1: How can we automatically detect defects/failures of chassis components (i.e, wheels and axle bearings) for train vehicles? From the technical perspective, is the trajectory-based maintenance solution feasible to be implemented in practical cases?

RQ2: How can we access and pre-processing the trajectory data in a proper way, such that it is capable to be incorporated in subsequent supervised machine leaning models?

RQ3: On the basis of RQ1 and RQ2, can we perform a proof-of-concept (PoC) demonstrator to provide the potential direction of application realisation and technology implementation for the feature?

5.4. Methodology and Tools

As we mentioned in the section of Objectives and Research Questions (Section 5.3), the primary task of this PoC is to explore whether the historical trajectory and train movement records can be used as a feasible solution to participate in the predictive maintenance of chassis of locomotives. And if so, to what extent we can get relevant insights and based on which potential policy suggestion can be provided for a more accurate predictive maintenance. The computer models under the framework must accurately represent how the components behave when varying levels of degradation are taken into account, while also supplying the normal and abnormal signal signatures captured by sensors mounted on the bogie components. In order to give the means for the quantitative characterisation of the component state and assist the creation of warning limits, these sensor signals must be pre-processed. It is also ideal if the models can forecast how the component state will deteriorate, providing an estimation of the RUL(Remaining Useful Lifetime), probability/type of the occurrence of different failures. This requirement applies to condition-based maintenance strategies, which effectively enable the optimisation of maintenance schedules.

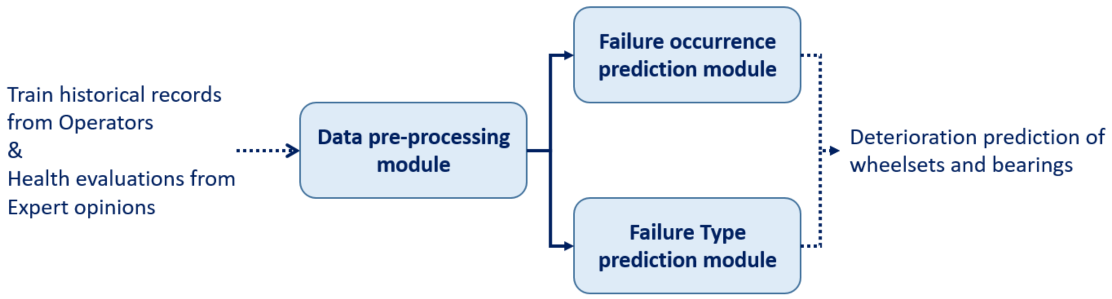

Fig. 5.2. High-Level Framework for CBM on chassis components

As we aim to achieve the prediction both the occurrence probability of axle bearing failures and what type/degree of this failure will be. The input data - Historical train movement records - can be collected and provided by the multiple train operators, while the latter - health evaluation opinions from experts, can be generated by those rolling stock maintenance teams on a fixed time basis. The outputs of the Smart Rolling Stock maintenance consists of two major modules/parts (As shown in figure 5.2). From our perspective, it is advisable to construct a diagram that illustrates data feeding, processing, modelling and predicting more explicitly.

Data pre-processing module is the procedure of processing raw data into more suitable characterisation inputs for subsequent modules.

Failure occurrence prediction module will produce the occurrence possibility of failures on a bogie component. Compare to the next module, its output more general and do not focus on any detailed anticipation since it only provide a percentage of how likely the failure on the components will happen in the future.

Failure type prediction module will provide the estimations of the health condition on more detailed level. Which corresponding to the 4-stage wheel/bearing wear we defined in section 5.2.

5.5. Artificial Intelligence and Machine Learning Models

Regarding the development of computational models for the forecasting of component deterioration [3], two different techniques can be distinguished:

- Making use of concepts from digital twins, we constructed prediction models based on extensive historical datasets of failures and observations on the same components of rolling stock. Such models are frequently ”black boxes,” (e.g. [55]) or completely mathematical abstractions, which do not employ any information on the physical processes of component deterioration and, as a result, should only be used as a last resort or a complementary technique for supporting other models, that is, when there is insufficient knowledge on the physics of the problem;
- Use of deterministic or stochastic models (e.g. [56]) that are dependent on expert knowledge and are an accurate representation of the mechanisms underlying degradation and failure, necessitating the description of the geometrical, material, and mechanical aspects that characterise the system These models' quality is constrained by the developer's skill and the availability of data to characterise the component's behaviour or failure mechanism.

In this case study we will give priority to the validation and implementation of ML-based models using historical train operational records and failures reports, not only because data driven models do not take into account the physics of the machinery but also because there is a more thorough and easily-obtained operational history database to support the use of artificial intelligence approaches. The digital twins-based models will be utilised to mimic the interaction between the vehicle and the track - Sensors are the essential part of the subsequent measurement, control, and diagnostic process. Device telemetry is collected using the smart sensors available on the hardware/software environment and then used to create the digital twin model of the physical equipment. All of the data is then aggregated and compiled to generate actionable information. The digital twin model is then continuously updated to mirror the current state of the physical thing. It can then be used to effectively model, monitor, and manage devices from a remote location. It also enables continuous intelligence & estimated time for the next needed maintenance, which the maintenance system can use to schedule at the optimal time. The outcomes will help create a database of the vehicle behaviour taking various levels of component deterioration into account.

The accuracy of this approach depends on the amount and quality of the data available. Computer simulations and the implementation of online condition monitoring have the potential to provide key support to extend the current inspection intervals. Computer simulations provide an estimate of the load spectra on the wheelset and bearings, that result from the vehicle-track interaction in the context of operation in specific lines/track segments and environmental conditions. The existing technologies for the condition monitoring of railway axles are still under development and are limited by the complexity of installation, data and power transmission, and the difficulty of the detection and characterisation of faults in such

[3] https://www.scribd.com/document/533802946/INNOWA-WP1-D-UNI-003-01-D1-1-Benchmark-and-Market-Drivers-for-an-Integrated-Intelligent-and-Lightweight-Wagon-Solution-1

a noisy environment. Online condition monitoring is not expected to replace the existing types of inspection in the foreseeable future but has the potential to complement the current inspection strategies.

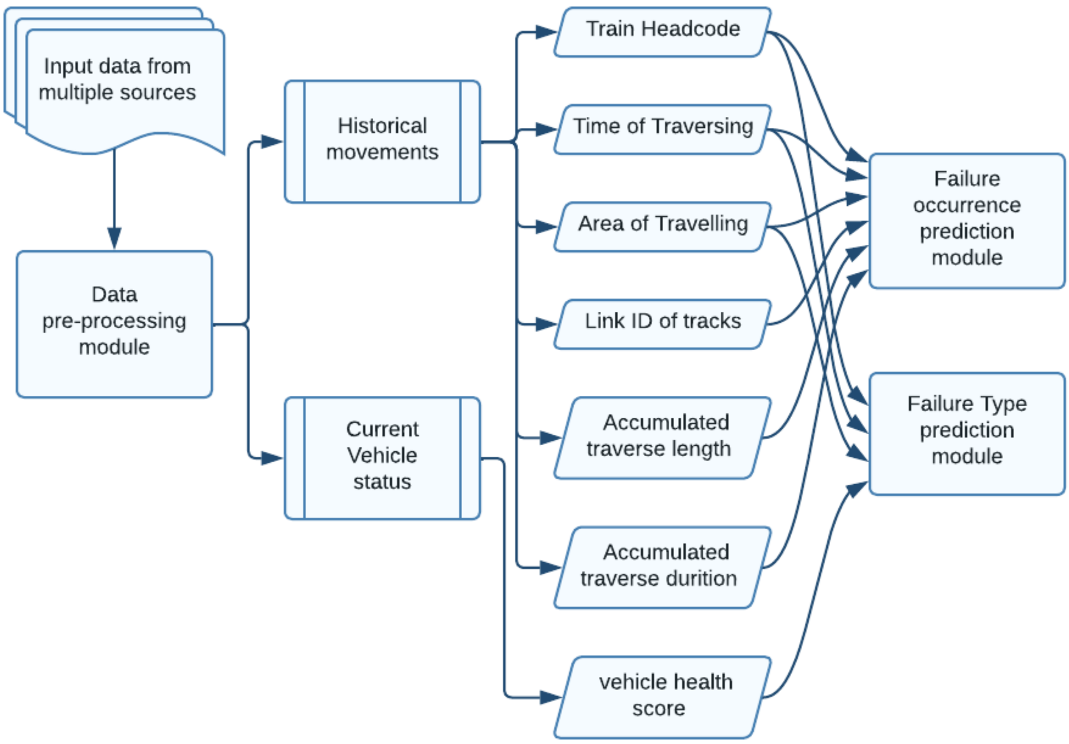

Fig. 5.3. The flow of data for Smart Rolling Stock Maintenance

To clearly illustrate the flow of data between the various modules and the specific set of features considered, we expand the macro-architecture in Figure 5.2 to the proposal we discussed above into Figure 5.3. As for the input features, we divided them into two categories based on their collecting manners. That is, one category we called as ”Historical Vehicle Movements Records”, which is extracted from multiple sources (i.e., operating logs/reports, infrastructure manager). Another category named as ”Current Rolling Stock Health Scores”, which is manually generated/evaluated based on the opinion of maintenance staff/team, which gives the information about the latest status of rolling stock health condition. In the former module that performing prediction of failure occurrence possibility, only Historical Vehicle Movements Records are leveraged as input. Notably, some of them can be directly collected or extracted from the existing data warefare (e.g., train operational logs and infrastructure provider reports). This module is typically a regression task and the expected output is the probability of failure of a component on a train service.

Table 5.2: Suitable AI approaches

Module	Contributions	Suitable Approaches
Failure Occurrence Prediction	High	Gradient Boosting Decision Tree
Failure Type Prediction	Medium	Random Forest Classifier, Multi-Layer Perceptron Classifier
Predictive Maintenance Scheduling	High	Particle Swarm Optimization

Thus it is reasonable to perform an Gradient Boosted Regression Tree model in this module. The Gradient boosting works by building simpler (weak) prediction models sequentially where each model runs to predict the error left over by the previous model (See Table 5.2). Because of this, the algorithm tends to converge rather quick. It is convinced that this model normally shown a better result in respect to traditional ML approaches or other conventional regression models [57] and [58]. In gradient boosting decision trees, we combine many weak learners to come up with one strong learner. The weak learners here are the individual decision trees. All the trees are connected in sequences and each tree tries to minimise the error of the previous tree. Due to this sequential connection, boosting algorithms are usually slow to learn, but also highly accurate. In statistical learning, models that learn slowly perform better. We give an example of a train service passing through a 'notorious' area where many very steep slopes on mountain-climbing or downhill tracks. The information such as Train Headcode, Time of Traversing this specific area, the area description of the vehicle traversed, the link identification of a particular segment of tracks, accumulated travelling length on the certain track section, and the time used on this section can be easily found in routine working records of train operators. While for the module of estimating the failure type, we additionally added the feature of vehicle evaluated health condition score into prediction. This will help to improve the performance of proposed condition-based maintenance framework since the main stream prediction paradigm on current train vehicle maintenance is based on knowledge base from experts. In this module, we propose to use several ML-based classifiers (i.e., Random forest Classifier[4], Multi-Layer Classifier[5]) to assess and evaluate the performance on 4-category classification task. To conclude, Table 5.2 summarises the information above and identifies the most promising candidate approaches for each separate module; In line with the previous study, column "contributions" gives a qualitative score (i.e., medium, or high) indicating the coverage of the corresponding module. It is worth to note that the first two modules are regarded as the implementation basis of the last module - the 'maintenance scheduling' module is the essential component that provides policy making suggestions to according to the output of the previous modules. And through literature investigation we suggested to apply a genetic algorithm (Particle Swarm Optimization).

[4] https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html

[5] https://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPClassifier.html

5.6. Reference Datasets

From the perspective of existing data collecting strategy, vibrations are the primary source of data collected by the sensors. To the gateway, each sensor transmits a full frequency spectrogram. A solar panel powers the gateway, which is a self-contained wireless gadget. It quickly captures all measurement data from the sensors across a Bluetooth low energy (BLE) network [59]. It uses GPS to track the vehicle (every per hour) and sends the entire data bundle to the cloud through widely available Long-Term Evolution (LTE) telecommunications. These distributed entities communicate over the network [60–62] using a diverse set of protocols and standards. The most common communication technologies for short range low power communication protocols are RFID (Radio Frequency Identification) [63] and NFC (Near Field Communication) [64]. As an additional data point, the temperature is collected. When an axle bearing overheats, it usually indicates that there has been a problem for some time, therefore the temperature data can establish that there is no overheating. This type of sensor is currently only available for freight trains, and train operators can pre-order them. The solution is promising to be modified to work with passenger trains as well. There are many commercial developers and open source offerings for providing middleware services to these IoT devices we introduced. Some examples are OpenIoT [65], MiddleWhere [66], Hydra [67], FiWare [68], etc.

However, in this section, we won't spend a lot of space on discussing many technical aspects of the sensor collected data, we will rather focus more on the train trajectory data synchronisation and how this data will be transmitted back to the computational center of predictive maintenance. The reduction in computation and sensor-free modelling has certainly helped the development of predictive maintenance on wheels and bearings, but there are still barriers to their widespread implementation. Important examples include the accessible for trajectory data and for a suitable AI-based option for developing predicting environment. There are a number of available train operating data resources can be used openly, for example, Darwin Data feeds [6] from National Rail data portal, and open data portal provided by Office of Rail and Road (ORR)[7]. These two open data sources provide sufficient and comprehensive information about how and when a vehicle started to traverse a specific area, and to what time it ends. Besides, it gives us more specifications regarding the running status of the rolling stock components after every journey.

Basing on what has been discussed so far, we summarize the key information as follow: it would be better to exploit real-world data instead of synthetic data collected from any of the simulators, however, there is no dataset directly exploitable has been found, only some online open data feeds and repositories are potentially available for building a suitable and qualified dataset.

5.7. Expected Results and Possible Criticalities

To sum up, this case study demonstrates a proof-of-concept about how digital twins approaches and machine learning-based models could be leveraged to deal with non-intrusive maintenance tasks for rolling stock components. Technically, our proposed method has unparalleled advantages in terms of implementation feasibility, for example, compared to

[6] https://www.nationalrail.co.uk/100296.aspx
[7] https://dataportal.orr.gov.uk/

traditional sensor-based methods, a large number of on-board or track-side sensors or cameras have to be installed for continuously capture the health condition information of monitored components. On the contrary, we can extract more useful information from the existing train operation data and leveraging them into the predictive maintenance tasks of the locomotive chassis components.

Data is essentially required to maintain an up-to-date digital twin throughout its entire operational life. A ML model must be paired with relevant data for it to become a digital twin. A model without situational data is generic, while with data, it is unique. Furthermore, collecting data just once is rarely enough. It must be timely and in agreement with the real-world system it is presenting [69]. As the cost of computing and communication continues to fall, the on-board IoT will become increasingly widespread and provide the necessary data flows more easily. These failure data streams collected by the train operators or mechanical built-in sensor are extremely important to support the functional of predictive maintenance procedure.

It is worth noting that the trajectory-based predictive maintenance framework we propose in this case study cannot completely replace sensor-based maintenance methods or periodic intrusive maintenance methods from the practical perspective. As an effective complement to traditional hardware-based methods, it is expected to provide more predictive insights for passive maintenance strategy before failures actually being detected. To this aim, we will give suggested probability of failure for a specific rolling stock component and the potential Remaining Useful Lifetime as the output of our paradigm, and such outputs can be evaluated and validated by the ground-truth health condition results that have been assessed/captured in real-life maintenance activities. Thereby, such paradigm is promising to make maintenance tasks for rolling stock chassis components more reliable, precise and economical.

Future work will begin with locating exportable/available datasets that might be relevant to our goals among those mentioned in this chapter. We will then create/collect small-scale mock datasets to adequately train the discussed modules using these datasets or additional sources (i.e., Failure occurrence prediction, Failure type prediction). Delineating roadmaps and guidelines for the health status evaluation for rail vehicle components especially rolling stock are thus our key future goals, commencing with the realization of the proposed proof-of-concept, or the realization of the suggested modules.

5.7.1. Opportunities for AI-enabled Digital Twins and Self-Protection Mechanisms

It is significant to highlight the potential of AI combined with Digital Twins (DTs) when it comes to smart maintenance applications, even though this is not the primary focus of our current research scope. AI could be one of the key boosters when developing DT components and it can be used in the many architectural levels that enable the DT application with various levels of automation. Without getting into any particular technical detail, our opinion is that AI might be involved in different ways when developing a DT: 1) It allows the safety and maintenance department to constantly monitor machines and collect data on the actual performance of a device against its expected work performance. The information can then be used to improve machine capability and maximize remaining useful lifetime. and 2) It provides a real-time connection between physical HW and analytical software, which empowers the process of aggregating data from every embedded machine into one location for more accessible analysis by the vehicle asset maintenance team.

It would be possible and necessary to (manually or automatically) perform some self-healing/self-adaptive mechanisms. It is able to provide a digital twin for each physical machinery component entity at the maintenance site. Intelligent software updates the twin's status in real-time to provide consolidated information on all aspects of the device. This information can be used by business and maintenance staff to quickly spot problems and identify the defects, based on which effective actions can be taken before they escalate into more significant issues that affect overall safety and operating efficiency. According to the taxonomies developed in [70], depending on the type of attack, self-healing attempts to determine the best reconfiguration or recovery action plan when those threats are detected. Simple faults can be addressed using legacy fault-tolerance mechanisms, which is a response policy based on redundancy, fault-isolation, and error correction within a single device; More complex cyber-physical attacks may need to adopt multi-step and multi-entity coordinated plans in practice. Thanks to 'what-if' analyses and predictions carried out in real-time by the run-time models inside DT, systems will be able to decide for themselves whether a shutdown or switchover to another external—most likely human-controlled—system is necessary or whether solutions are available to counter threats and restore system operation, possibly with reduced capacity or performance.

The adaptive digital twin provides the physical and digital twins with an adaptive user interface. The adaptive user interface takes the user or operator's preferences and priorities into consideration. Learning human operators' preferences and priorities in various settings is a crucial skill at this level [71]. An approach for supervised machine learning built on neural networks captures these traits. Based on the data that is "drawn" in real-time from the physical twin, the models used within this digital twin are continuously updated. Following system use, it can also take data in batches. During operations, maintenance, and support, this digital twin can aid in real-time planning and decision-making. Furthermore, a solution based in DT can be incorporated into a data usage platform that makes it easy to bring together and prepare OT/IT data saved individually by many different operations. It helps make continuous maintenance efficiency improvements through AI analysis and simulation. This kind of platform uses the links between maintenance operations to model site data.

6. Conclusions

In this deliverable, we have addressed the definition of two relevant case studies in the context of AI for smart and predictive maintenance and defect detection in railways. The case studies will be developed in the continuation of the project in order to get some useful insights to support the definition of future roadmaps in the use of AI approaches in related railway applications. At this stage, the case studies have been specified mainly in terms of background and motivation, objectives and research questions, planned AI methodologies and tools, available datasets, AI/ML models to be used, expected results and possible criticalities. We expect that actual experimentation during the proof-of-concept stage will further highlight a set of opportunities and challenges to be addressed by researchers and engineers in order to leverage on the full potential of AI and ML.

Due to the project scope, a selection of relevant scenarios had to be made considering the planned efforts, limited time requirements, and results achievable; however, we highlighted a comprehensive set of research directions and perspectives that could be explored in addition to the two case-studies described in this deliverable.

Bibliography

[1] R. Tang, L. De Donato, N. Besinović, F. Flammini, R. M. Goverde, Z. Lin, R. Liu, T. Tang, V. Vittorini, and Z. Wang, "A literature review of artificial intelligence applications in railway systems," *Transportation Research Part C: Emerging Technologies*, vol. 140, p. 103679, 2022.

[2] RAILS, "Deliverable 3.1 – WP3 Report on case studies and analysis of transferability from other sectors (predictive maintenance and defect detection)," 2021. [Online]. Available: https://rails-project.eu/downloads/deliverables

[3] European Union Agency for Railways (ERA), "Safety overview 2021," Tech. Rep., 2021.

[4] ——, "Report on Railway Safety and Interoperability in the EU," Tech. Rep., 2020.

[5] SAFER-LC, "SAFER Level Crossing by integrating and optimizing road-rail infrastructure management and design," 2017-2020. [Online]. Available: https://safer-lc.eu/

[6] P. Singh, M. A. Dulebenets, J. Pasha, E. D. S. Gonzalez, Y.-Y. Lau, and R. Kampmann, "Deployment of autonomous trains in rail transportation: Current trends and existing challenges," *IEEE Access*, vol. 9, pp. 91 427–91 461, 2021.

[7] M. A. B. Fayyaz, A. C. Alexoulis-Chrysovergis, M. J. Southgate, and C. Johnson, "A review of the technological developments for interlocking at level crossing," *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, vol. 235, no. 4, pp. 529–539, 2021.

[8] O. of Rail Regulation, "Level crossings: A guide for managers, designers and operators," 2011. [Online]. Available: https://www.orr.gov.uk/sites/default/files/om/level_crossings_guidance.pdf

[9] N. Rail, "Nr/l2/sig/19608 - level crossing asset inspection and implementation of minimum action codes," 2014. [Online]. Available: https://standards.globalspec.com/std/1689019/nr-l2-sig-19608

[10] ——, "Nr/l2/sig/19608 issue 8 - inspection of level crossing systems," 2021. [Online]. Available: https://standards.globalspec.com/std/14461410/nr-l2-xng-19608-issue-8

[11] R. Dirnfeld, L. De Donato, F. Flammini, M. S. Azari, and V. Vittorini, "Railway digital twins and artificial intelligence: Challenges and design guidelines," in *European Dependable Computing Conference*. Springer, 2022, pp. 102–113.

[12] SAFER-LC, "Deliverable d1.3: Needs and requirements for improving level crossing safety," 2018.

[13] M. Martinez, C. Sitawarin, K. Finch, L. Meincke, A. Yablonski, and A. Kornhauser, "Beyond grand theft auto v for training, testing and enhancing deep learning in self driving cars," *arXiv preprint arXiv:1712.01397*, 2017.

[14] B. Hurl, K. Czarnecki, and S. Waslander, "Precise synthetic image and LiDAR (presil) dataset for autonomous vehicle perception," in *2019 IEEE Intelligent Vehicles Symposium (IV)*, 2019, pp. 2522–2529.

[15] A. Karovalia. (2020) Building a self-driving vehicle in GTA-V using deep learning and convolutional neural network. Accessed: 24 May 2022. [Online]. Available:

https://alzaibkarovalia.medium.com/building-a-self-driving-vehicle-in-gta-v-using-deep-learning-and-convolutional-neural-network-696b38b4c81e

[16] M. J. Pappaterra, F. Flammini, V. Vittorini, and N. Bešinović, "A systematic review of artificial intelligence public datasets for railway applications," *Infrastructures*, vol. 6, no. 10, p. 136, 2021.

[17] S. Hershey, S. Chaudhuri, D. P. Ellis, J. F. Gemmeke, A. Jansen, R. C. Moore, M. Plakal, D. Platt, R. A. Saurous, B. Seybold *et al.*, "CNN architectures for large-scale audio classification," in *Int. Conf. on Acoustics, Speech and Signal Processing*. IEEE, 2017, pp. 131–135.

[18] J. F. Gemmeke, D. P. Ellis, D. Freedman, A. Jansen, W. Lawrence, R. C. Moore, M. Plakal, and M. Ritter, "Audio set: An ontology and human-labeled dataset for audio events," in *Int. Conf. on Acoustics, Speech and Signal*. IEEE, 2017, pp. 776–780.

[19] L. De Donato, F. Flammini, S. Marrone, C. Mazzariello, R. Nardone, C. Sansone, and V. Vittorini, "A survey on audio-video based defect detection through deep learning in railway maintenance," *IEEE Access*, pp. 1–1, 2022.

[20] R. Tang, L. De Donato, N. Besinović, F. Flammini, R. M. Goverde, Z. Lin, R. Liu, T. Tang, V. Vittorini, and Z. Wang, "A literature review of artificial intelligence applications in railway systems," *Transportation Research Part C: Emerging Technologies*, vol. 140, p. 103679, 2022.

[21] RAILS, "Deliverable D1.2: Summary of existing relevant projects and state-of-the-art of AI applications in railways," 2021, DOI: 10.13140/RG.2.2.11353.03686. [Online]. Available: https://rails-project.eu/downloads/deliverables

[22] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on knowledge and data engineering*, vol. 22, no. 10, pp. 1345–1359, 2009.

[23] R. Kulkarni, S. Dhavalikar, and S. Bangar, "Traffic light detection and recognition for self driving cars using deep learning," in *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*, 2018, pp. 1–4.

[24] Q. Qi, F. Tao, T. Hu, N. Anwer, A. Liu, Y. Wei, L. Wang, and A. Nee, "Enabling technologies and tools for digital twin," *Journal of Manufacturing Systems*, vol. 58, pp. 3–21, 2021.

[25] F. Flammini, L. De Donato, A. Fantechi, and V. Vittorini, "A vision of intelligent train control," in *International Conference on Reliability, Safety, and Security of Railway Systems*. Springer, 2022, pp. 192–208.

[26] RAILS, "Deliverable D2.1: WP2 Report on case studies and analysis of transferability from other sectors (safety and automation)," 2021. [Online]. Available: https://rails-project.eu/downloads/deliverables

[27] D. Orive, N. Iriondo, A. Burgos, I. Saráchaga, M. L. Álvarez, and M. Marcos, "Fault injection in digital twin as a means to test the response to process faults at virtual commissioning," in *2019 24th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, 2019, pp. 1230–1234.

[28] R. Pfaff, P. Shahidi, and M. Enning, "Connected freight rail rolling stock: a modular approach integrating sensors, actors and cyber physical systems for operational advantages and condition based maintenance," in *Proceedings of the Asia Pacific Conference of the Prognostics and Health Management Society*, 2017.

[29] M. Papaelias, A. Amini, Z. Huang, P. Vallely, D. C. Dias, and S. Kerkyras, “Online condition monitoring of rolling stock wheels and axle bearings,” *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, vol. 230, no. 3, pp. 709–723, 2016.

[30] C. Sriskandarajah, A. K. Jardine, and C. Chan, “Maintenance scheduling of rolling stock using a genetic algorithm,” *Journal of the Operational Research Society*, vol. 49, no. 11, pp. 1130–1145, 1998.

[31] M. Penicka, A. K. Strupchanska, and D. Bjørner, “Train maintenance routing,” in *FORMS’2003: Symposium on Formal Methods for Railway Operation and Control Systems*, 2003.

[32] G. Maróti and L. Kroon, “Maintenance routing for train units: the transition model,” *Transportation Science*, vol. 39, no. 4, pp. 518–525, 2005.

[33] G. Budai, D. Huisman, and R. Dekker, “Scheduling preventive railway maintenance activities,” *Journal of the Operational Research Society*, vol. 57, no. 9, pp. 1035–1044, 2006.

[34] L. Wang, J. Chu, and J. Wu, “Selection of optimum maintenance strategies based on a fuzzy analytic hierarchy process,” *International journal of production economics*, vol. 107, no. 1, pp. 151–163, 2007.

[35] T. Delbrügger, M. Meißner, A. Wirtz, D. Biermann, J. Myrzik, J. Rossmann, and P. Wiederkehr, “Multi-level simulation concept for multidisciplinary analysis and optimization of production systems,” *The International Journal of Advanced Manufacturing Technology*, vol. 103, no. 9, pp. 3993–4012, 2019.

[36] D. Li, X. Yang, and X. Xu, *A Framework of Smart Railway Passenger Station Based on Digital Twin*, pp. 2623–2634. [Online]. Available: https://ascelibrary.org/doi/abs/10.1061/9780784482933.226

[37] M. Hong, Q. Wang, Z. Su, and L. Cheng, “In situ health monitoring for bogie systems of crh380 train on beijing–shanghai high-speed railway,” *Mechanical Systems and Signal Processing*, vol. 45, no. 2, pp. 378–395, 2014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0888327013006444

[38] J. A. Silmon, “Operational industrial fault detection and diagnosis: railway actuator case studies,” Ph.D. dissertation, University of Birmingham, 2009.

[39] B. P. Dahmer *et al.*, “Bearing maintenance practices to ensure maximum bearing life,” in *Proceedings of the 29th International Pump Users Symposium*. Turbomachinery Laboratories, Texas A&M Engineering Experiment Station, 2013.

[40] T. Otte, A. F. Posada-Moreno, F. Hübenthal, M. Haßler, H. Bartels, A. Abdelrazeq, and F. Hees, “Condition monitoring of rail infrastructure and rolling stock using acceleration sensor data of on-rail freight wagons.” in *ICPRAM*, 2022, pp. 432–439.

[41] C. A. Avizzano, G. Scivoletto, and P. Tripicchio, “Robust image stitching and reconstruction of rolling stocks using a novel kalman filter with a multiple-hypothesis measurement model,” *IEEE Access*, vol. 9, pp. 154 011–154 021, 2021.

[42] B. P. Graney and K. Starry, “Rolling element bearing analysis,” *Materials Evaluation*, vol. 70, no. 1, p. 78, 2012.

[43] M. Papaelias, Z. Huang, A. Amini, P. Vallely, N. Day, R. Sharma, Y. Kerkyras, and

S. Kerkyras, “Advanced wayside condition monitoring of rolling stock wheelsets,” *the Proceedings of the 11 th ECNDT, Prague, Czech Republic*, pp. 6–10, 2014.

[44] R. Lagnebäck, “Evaluation of wayside condition monitoring technologies for condition-based maintenance of railway vehicles,” Ph.D. dissertation, Luleå tekniska universitet, 2007.

[45] A. Golubenko, S. Sapronova, and V. Tkacenko, “Kinematics of point-to-point contact of wheels with a rails,” *Transport problems*, vol. 2, no. 3, pp. 57–61, 2007.

[46] L. Smetanka, P. Št'astniak, and J. Harušinec, “Wear research of railway wheelset profile by using computer simulation,” in *MATEC Web of conferences*, vol. 157. EDP Sciences, 2018, p. 03017.

[47] G. VAičiūnAs, G. BureikA, and S. Steišūnas, “Research on metal fatigue of rail vehicle wheel considering the wear intensity of rolling surface,” *Eksploatacja i Niezawodność*, vol. 20, no. 1, 2018.

[48] V. Tkachenko and S. Sapronova, “Steerability of railway vehicles,” *Transport problems*, vol. 2, no. 4, pp. 9–16, 2007.

[49] A. Falendysh, P. Kharlamov, O. Kletska, and N. Volodarets, “Calculation of the parameters of hybrid shunting locomotive,” *Transportation Research Procedia*, vol. 14, pp. 665–671, 2016.

[50] J. Gerlici, T. Lack, and J. Harušinec, “Realistic simulation of railway operation on the railbcot test stand,” in *Applied mechanics and materials*, vol. 486. Trans Tech Publ, 2014, pp. 387–395.

[51] J. Gerlici, T. Lack, and J. Harusinec, “Development of test stand prototype for rail vehicles brake components testing,” *Communications-Scientific letters of the University of Zilina*, vol. 16, no. 3A, pp. 27–32, 2014.

[52] A. Sapietová and V. Dekỳš, “Dynamic analysis of rotating machines in msc. adams,” *Procedia Engineering*, vol. 136, pp. 143–149, 2016.

[53] M. Chalupa, L. Severa, and R. Vlach, “Vehicle track loading simulation,” in *Fourth International Conference on Machine Vision (ICMV 2011): Computer Vision and Image Analysis; Pattern Recognition and Basic Technologies*, vol. 8350. SPIE, 2012, pp. 170–174.

[54] M. Chalupa, J. Veverka, and R. Vlach, “Dynamic loading of vehicle track simulation,” *Engineering Mechanics*, 2011.

[55] K. T. Nguyen and K. Medjaher, “A new dynamic predictive maintenance framework using deep learning for failure prognostics,” *Reliability Engineering & System Safety*, vol. 188, pp. 251–262, 2019.

[56] A. Ray and S. Tangirala, “Stochastic modeling of fatigue crack dynamics for on-line failure prognostics,” *IEEE Transactions on Control Systems Technology*, vol. 4, no. 4, pp. 443–451, 1996.

[57] P. Bahad and P. Saxena, “Study of adaboost and gradient boosting algorithms for predictive analytics,” in *International Conference on Intelligent Computing and Smart Communication 2019*. Springer, 2020, pp. 235–244.

[58] P. Wang and Q.-p. Zhang, “Train delay analysis and prediction based on big data fusion,” *Transportation Safety and Environment*, vol. 1, no. 1, pp. 79–88, 2019.

[59] K.-H. Chang, "Bluetooth: a viable solution for iot?[industry perspectives]," *IEEE Wireless Communications*, vol. 21, no. 6, pp. 6–7, 2014.

[60] L. Atzori, A. Iera, and G. Morabito, "The internet of things: A survey," *Computer networks*, vol. 54, no. 15, pp. 2787–2805, 2010.

[61] A. Whitmore, A. Agarwal, and L. Da Xu, "The internet of things—a survey of topics and trends," *Information systems frontiers*, vol. 17, no. 2, pp. 261–274, 2015.

[62] D. Zeng, S. Guo, and Z. Cheng, "The web of things: A survey," *J. Commun.*, vol. 6, no. 6, pp. 424–438, 2011.

[63] R. Want, "An introduction to rfid technology," *IEEE pervasive computing*, vol. 5, no. 1, pp. 25–33, 2006.

[64] ——, "Near field communication," *IEEE Pervasive Computing*, vol. 10, no. 3, pp. 4–7, 2011.

[65] J. Soldatos, N. Kefalakis, M. Hauswirth, M. Serrano, J.-P. Calbimonte, M. Riahi, K. Aberer, P. P. Jayaraman, A. Zaslavsky, I. P. Žarko *et al.*, "Openiot: Open source internet-of-things in the cloud," in *Interoperability and open-source solutions for the internet of things*. Springer, 2015, pp. 13–25.

[66] A. Ranganathan, J. Al-Muhtadi, S. Chetan, R. Campbell, and M. D. Mickunas, "Middlewhere: a middleware for location awareness in ubiquitous computing applications," in *ACM/IFIP/USENIX International Conference on Distributed Systems Platforms and Open Distributed Processing*. Springer, 2004, pp. 397–416.

[67] M. Eisenhauer, P. Rosengren, and P. Antolin, "A development platform for integrating wireless devices and sensors into ambient intelligence systems," in *2009 6th IEEE annual communications society conference on sensor, mesh and ad hoc communications and networks workshops*. IEEE, 2009, pp. 1–3.

[68] T. Zahariadis, A. Papadakis, F. Alvarez, J. Gonzalez, F. Lopez, F. Facca, and Y. Al-Hazmi, "Fiware lab: managing resources and services in a cloud federation supporting future internet applications," in *2014 IEEE/ACM 7th International Conference on Utility and Cloud Computing*. IEEE, 2014, pp. 792–799.

[69] G. C. Doubell, K. Kruger, A. H. Basson, and P. Conradie, "The potential for digital twin applications in railway infrastructure management," in *World Congress on Engineering Asset Management*. Springer, 2021, pp. 241–249.

[70] A. Avizienis, J.-C. Laprie, B. Randell, and C. Landwehr, "Basic concepts and taxonomy of dependable and secure computing," *IEEE transactions on dependable and secure computing*, vol. 1, no. 1, pp. 11–33, 2004.

[71] A. M. Madni, M. G. Samet, and A. Freedy, "A trainable on-line model of the human operator in information acquisition tasks," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 12, no. 4, pp. 504–511, 1982.